

WHAT IS CLAIMED IS:

1. A method for identifying an XML document, comprising the steps of:

obtaining a document;

matching the document against a plurality of XML schemas that specify a

5 set of document types; and

based on the result of the matching step, outputting information regarding
the document.

2. The method of claim 1, wherein the outputted information includes information

10 regarding the identity of the document type.

3. The method of claim 1, wherein the matching step includes determining match
scores.

15 4. The method of claim 3, wherein each of the match scores reflects the degree of
closeness between the document and one of the XML schemas.

5. The method of claim 4, wherein a match score of zero indicates a perfect match.

20 6. The method of claim 4, wherein a non-zero match score indicates a mismatch.

7. The method of claim 3, wherein determining the match scores includes
determining the match scores by performing minimum-mismatch comparisons.

8. The method of claim 1, wherein the document is received from an external source.

9. The method of claim 8, wherein the external source uses the outputted
5 information to perform a categorization process before performing further operations on the document.

10. The method of claim 8, wherein the external source uses the outputted information to route the document.

10

11. The method of claim 8, wherein the external source uses the outputted information to determine whether the document passes a first-level validation.

12. The method of claim 1, wherein the document is undergoing incremental change.

15

13. The method of claim 1, wherein the outputted information includes confirmation that the document conforms to a known document structure.

20

14. A system for identifying an XML document, comprising:
an input component for obtaining a document;
a validation component for matching the document against a plurality of XML
schemas that specify a set of document types; and
5 an output component for outputting information regarding the document
indicating the results of the matching.

15. The system of claim 14, wherein the outputted information includes information
regarding the identity of the document type.

10

16. The system of claim 14, wherein the validation component determines match
scores.

15

17. The system of claim 16, wherein each of the match scores reflects the degree of
closeness between the document and one of the XML schemas.

18. The system of claim 17, wherein a match score of zero indicates a perfect match.

19. The system of claim 17, wherein a non-zero match score indicates a mismatch.

20

20. The system of claim 16, wherein the validation component determines the match
scores by performing minimum-mismatch comparisons.

21. The system of claim 14, wherein the input component receives the document from an external source.

22 The system of claim 21, wherein the external source uses the outputted
5 information to perform a categorization process before performing further operations on the document.

23. The system of claim 21, wherein the external source uses the outputted information to route the document.

10

24. The system of claim 21, wherein the external source uses the outputted information to determine whether the document passes a first-level validation.

25. The system of claim 14, wherein the document is undergoing incremental change.

15

26. The system of claim 14, wherein the outputted information includes confirmation that the document conforms to a known document structure.

20

27. A program storage device readable by a machine, tangibly embodying a program of instructions executable on the machine to perform method steps for identifying an XML document , the method steps comprising:

obtaining a document;

5 matching the document against a plurality of XML schemas that specify a set of document types; and

based on the result of the matching step, outputting information regarding the document.

10

15